

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-101203

(P2002-101203A)

(43)公開日 平成14年4月5日(2002.4.5)

(51)Int.Cl. ⁷	識別記号	F I	テ-マ-ト*(参考)
H 0 4 M 3/42		H 0 4 M 3/42	P 5 D 0 1 5
G 1 0 L 13/00		1/00	H 5 D 0 4 5
15/00		3/50	A 5 K 0 1 5
15/22		11/00	3 0 2 5 K 0 2 4
15/28		G 1 0 L 3/00	R 5 K 0 2 7
審査請求 未請求 請求項の数6 OL (全 5 頁) 最終頁に続く			

(21)出願番号 特願2000-285838(P2000-285838)

(22)出願日 平成12年9月20日(2000.9.20)

(71)出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72)発明者 室井 哲也

東京都大田区中馬込1丁目3番6号 株式会社リコー内

Fターム(参考) 5D015 KK01 LL06

5D045 AA07 AB24 AB26

5K015 AA07

5K024 BB01 BB04 CC01 EE09

5K027 AA10 HH20

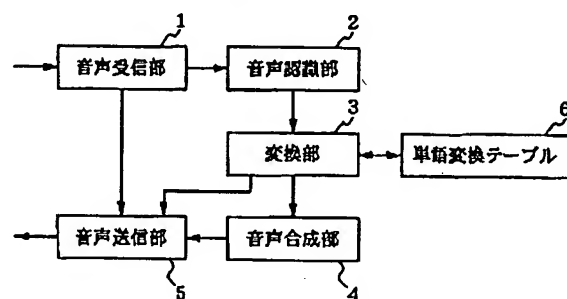
5K101 LL02 NN08

(54)【発明の名称】 音声処理システム、音声処理方法およびその方法を記憶した記憶媒体

(57)【要約】

【課題】 話者の特有な言い回しや独特な表現方法、口癖など様々な特徴を含む言葉を特徴のない一般的な言葉に変換することにより、話者が誰であるかわからなくできる音声処理システムを提供する。

【解決手段】 通信手段を介して音声信号を通信する音声処理システムにおいて、話者の発した音声信号について音声認識を行う音声認識部2、変換前の単語と変換後の単語の対を記憶しておく単語変換テーブル6、音声認識部2により認識された単語が単語変換テーブルの変換前の単語として登録されていた場合に対応する変換後の単語に変換する変換部3、変換された単語から合成音声信号を生成する音声合成部4を備え、音声合成部4により合成音声信号が生成されたとき、音声送信部5が、登録されていた前記変換前の単語の音声信号を変換後の単語の音声合成信号に置換する構成にした。



(2)

1

【特許請求の範囲】

【請求項1】 通信手段を介して音声信号または音声データを通信する音声処理システムにおいて、話者の発した音声信号について音声認識を行う音声認識手段と、前記音声認識手段により認識された少なくとも一部の音声データを記憶されている音声データに変換する変換手段と、前記変換手段により変換された音声データから合成音声信号を生成する音声合成手段と、前記音声信号中の対応する部分を前記音声合成手段により生成された合成音声信号に置換する音声置換手段とを備えたことを特徴とする音声処理システム。

【請求項2】 請求項1記載の音声処理システムにおいて、変換前の単語と変換後の単語の対を記憶しておく単語変換テーブルを備え、音声認識手段により認識された音声信号中の単語が前記単語変換テーブルに変換前の単語として登録されていた場合に、登録されていた前記単語を対応する変換後の単語に変換する構成にしたことを特徴とする音声処理システム。

【請求項3】 請求項1または請求項2記載の音声処理システムにおいて、置換された合成音声信号を含む音声信号を公衆電話回線へ送出する公衆回線通信手段を備えたことを特徴とする音声処理システム。

【請求項4】 請求項1または請求項2記載の音声処理システムにおいて、置換された合成音声信号を含む音声信号をデータ通信ネットワークへ送出するデータ通信手段を備えたことを特徴とする音声処理システム。

【請求項5】 通信手段を介して音声信号または音声データを通信する音声処理方法において、話者の発した音声信号について音声認識を行い、認識された音声データが登録されていたならば、その音声データを対応づけて記憶されている他の音声データに変換し、変換された音声データから合成音声信号を生成し、前記音声信号中の対応する部分を生成された合成音声信号に置換することを特徴とする音声処理方法。

【請求項6】 プログラムを記憶した記憶媒体において、請求項5記載の音声処理方法を実施するためのプログラムを記憶したことを特徴とする記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、ネットワークを介した会議や電話や音声メッセージ討論など複数の話者が対話する環境を提供する音声処理システムに係わり、特に、話者が誰であるかを秘匿にすることができる音声処理システムに関する。

【0002】

【従来の技術】 コンピュータネットワークの普及により、ネットワーク上で複数の話者が会話することにより作業を進めたり、会議をしたり、生活情報を交換したりすることができる音声対話システムが普及しつつある。また、このようなネットワークを用いた音声対話システ

2

ムにおいて、複数の話者が対話する場における対話に係わるサービスのひとつとして、話者が誰であるかを秘匿にすることができる音声対話システムが従来より提供されている。例えば、特開平9-83655号公報に示された音声対話システムは、このような音声対話システムのひとつであり、音声を音声信号に変換する音声入力手段、および音声信号を音声に変換する音声出力手段を備えるとともに、通信回線に接続される複数の端末装置と、通信回線を介してこれら複数の端末装置と接続され、その端末装置との間で音声信号の収集、配信を行なうサーバとを備え、さらに、音声をエフェクタを通すことにより音響的な特徴を変化させて話者を特定できないようにする。

【0003】

【発明が解決しようとする課題】 しかしながら、実際に人間が話者を特定する場合、ピッチ周波数（声帯の振動周波数）など音響的な特徴だけから特定するのではない。話者の特有な言い回しや独特な表現方法、口癖なども含む様々な特徴を利用して話者が誰であるかを特定するのである。したがって、特開平9-83655号公報に示されているようにエフェクタにより周波数軸上の変換を行なうだけでは、例えばポーズ時間、呼吸段落の割合や長さ、単語の継続時間など継続時間的特徴、つまり時間軸上の特徴や、話者の特有な言い回しや独特な表現方法、口癖なども含む様々な特徴から、話者が誰であるか容易に推定されてしまう場合がある。本発明の目的は、このような従来技術の問題を解決し、話者の特有な言い回しや独特な表現方法、口癖など様々な特徴を含む言葉の特徴のない一般的な言葉に変換することにより、話者が誰であるかわからなくできる音声処理システムを提供することにある。

【0004】

【課題を解決するための手段】 前記の課題を解決するために、請求項1記載の発明では、通信手段を介して音声信号または音声データを通信する音声処理システムにおいて、話者の発した音声信号について音声認識を行う音声認識手段と、前記音声認識手段により認識された少なくとも一部の音声データを記憶されている音声データに変換する変換手段と、前記変換手段により変換された音声データから合成音声信号を生成する音声合成手段と、前記音声信号中の対応する部分を前記音声合成手段により生成された合成音声信号に置換する音声置換手段とを備えた。また、請求項2記載の発明では、請求項1記載の発明において、変換前の単語と変換後の単語の対を記憶しておく単語変換テーブルを備え、音声認識手段により認識された音声信号中の単語が前記単語変換テーブルに変換前の単語として登録されていた場合に、登録されていた前記単語を対応する変換後の単語に変換する構成にした。また、請求項3記載の発明では、請求項1または請求項2記載の発明において、置換された合成音声信

(3)

3

号を含む音声信号を公衆電話回線へ送出する公衆回線通信手段を備えた。また、請求項4記載の発明では、請求項1または請求項2記載の発明において、置換された合成音声信号を含む音声信号をデータ通信ネットワークへ送出するデータ通信手段を備えた。また、請求項5記載の発明では、通信手段を介して音声信号または音声データを通信する音声処理方法において、話者の発した音声信号について音声認識を行い、認識された音声データが登録されていたならば、その音声データに対応づけて記憶されている他の音声データに変換し、変換された音声データから合成音声信号を生成し、前記音声信号中の対応する部分を生成された合成音声信号に置換する方法にした。また、請求項6記載の発明では、プログラムを記憶した記憶媒体において、請求項5記載の音声処理方法を実施するためのプログラムを記憶した。

【0005】

【発明の実施の形態】以下、図面により本発明の実施の形態を詳細に説明する。図1は本発明の第1の実施例を示す音声処理システムの構成ブロック図である。図示したように、この実施例の音声処理システムは、話者が発生した音声信号を公衆電話回線を介して受信する音声受信部1、その音声受信部1により受信された音声信号について音声認識を行う音声認識部2、その音声認識部2により認識された少なくとも一部の受信音声データ（文字列）を記憶されている対応する音声データに変換する変換部3、変換された音声データから合成音声信号を生成する音声合成部4、受信された音声信号の一部を合成音声信号に置換した音声信号を公衆電話回線を介して他の話者へ送信する音声送信部5、変換前の単語と変換後の単語の対を記憶しておく単語変換テーブル6を備える。なお、この実施例では、請求項記載の音声認識手段、変換手段、音声合成手段、音声置換手段は、それぞれその順に、音声認識部2、変換部3、音声合成部4、音声送信部5により実現され、公衆回線通信手段は音声受信部1および音声送信部5により実現される。なお、この音声処理システムは、データやプログラムを一時的に記憶するメモリ、そのプログラムに従って動作するCPU、データやプログラムを記憶しておくハードディスク装置を備え、前記変換部3はCPUやメモリから構成され、音声受信部1および音声送信部5はCPU、メモリ、および専用回路から構成され、音声認識部2および音声合成部4はCPU、メモリ、専用回路、およびハードディスク装置などから構成される。図2に、第1の実施例の動作フローを示す。以下、図2などに従って、この実施例の動作を説明する。この実施例の音声処理システムでは、まず対話に先立って、対話に参加するすべての電話機などが公衆電話回線を介してこの音声処理システムに接続される。例えば、ひとつの電話機からこの音声処理システムに発呼してその電話機とこの音声処理システムとの間に回線が接続された後、その電話機からブ

4

ッシュボタンによりひとつまたは複数の対話相手の電話番号を指定させる。これにより、その電話機から相手先電話番号を示すDTMF信号が送られてくると、音声受信部1内のDTMF信号検出手段がその相手先電話番号を検出し、検出した電話番号を音声送信部5に渡す。そうすると、音声送信部5がその電話番号の相手先に発呼して回線を接続させるのである。続いて、この音声処理システムに回線が接続されている複数の電話機中のひとつから話者秘匿を示すDTMF信号に続いて音声信号が送られてくると、音声受信部1がその音声信号を受信して先頭から順に音声認識部2に渡す。そうすると、音声認識部2は、ハードディスク装置内の構文辞書や単語辞書などを参照して当業者には公知の方法により文（音声）の冒頭から順に単語（または単語+助詞）を切り出して取得し（S1）、取得した単語について音声認識を行う。そして、音声認識部2により認識された単語Aを変換部3が取得すると、変換部3は取得した単語Aを単語変換テーブルの変換前の複数の特定単語と照合し、単語Aと同じ文字列の特定単語が登録されているか否かを調べる（S2）。

【0006】図3に、単語変換テーブルを示す。図3において、左欄は変換前の単語で、例えば話者特有の言い回しや独特な表現方法、口癖などを含む特定単語である。また、右欄は変換後の単語で一般的な表現方法が対応づけられている。なお、図3の例で、「シーブラブラ」に対応づけられた「シープラスプラス」とは「C+」（C言語の改良版）のことである。その結果、一致する変換前の単語Aが登録されていないと判定されたならば（S2でno）、変換部3は単語Aの置換を行なわない旨を音声送信部5に通知し、その通知を受けた音声送信部5はその部分の単語の置換を行なうことなく、音声受信部1から渡されたその部分の音声信号をそのまま話者の電話機以外の電話機（話者の電話番号は音声受信部1から取得することにより知る）へ送出する（S3）。それに対して、取得された単語Aに一致する変換前の単語が登録されていると判定されたならば（S2でyes）、変換部3は単語Aに対応づけて（対として）登録されている変換後の単語Bを取得し、その単語を音声合成部4に渡す。これにより、音声合成部4はハードディスク装置などに予め記憶してある所定のモデル情報に従って、当業者には公知の方法により渡された変換後の単語であるかな文字に対応した合成音声信号を生成し、それを音声送信部5に渡す。そうすると、音声送信部5は受信音声信号中の、渡された合成音声信号に対応した部分をその合成音声信号に置換し、置換された音声信号をステップS3の場合と同様に出力する（S4）。例えば、「私はシーブラブラには慣れています」という受信音声中の「シーブラブラ」を「シープラスプラス」に置換して出力するのである。取り出された1単語分（あるいは1単語+助詞）の音声送出手続きが終了すると（こ

(4)

5

の時点ではこの分の送出は始まったばかりである)、音声認識部2は文末(一連の音声信号の最後)まで達したかどうかを判定し(S5)、達していないと判定されたならば(S5でno)、受信音声信号中から次の単語を取り出し(S6)、ステップS2から繰り返す。そして、ステップS5において、文末に達したと判定されたならば(S5でyes)、この動作フローを終了させる。また、この実施例では、ポーズ時間や呼気段落の長さ、単語(または単語+助詞)または音節の継続時間などをこの音声処理システム内にあるタイマを用いて測定することにより、予め設定した所定の時間との誤差を求め、誤差が所定値以上であれば、予め設定した所定の時間に置換するようにする。なお、単語(助詞を含む)の継続時間については設定する所定の時間を音節の数から求められるようにしておく。こうして、この実施例によれば、話者の特有な言い回しや独特な表現方法、口癖などが一般的な言葉や表現に置換されるので、話者が誰であるかわからなくできる。

【0007】図4は、本発明の第2の実施例を示す構成ブロック図である。この実施例の音声処理システムは例えばデータ通信ネットワークに接続されたパーソナルコンピュータなど端末装置内に実施され、図示したように、図1に示した音声受信部1および音声送信部5の代わりに、他の端末装置(クライアント装置)やサーバとの間でデータ通信を行なうデータ通信制御部7を備え、さらに、マイクروفोनなどを有した音声入力部8、スピーカなどを有した音声出力部9などを備える。なお、請求項記載のデータ通信手段は、この実施例では、データ通信制御部7により実現される。このような構成で、この実施例の音声処理システムでは、例えば、音声入力部6により入力した音声信号について音声認識部2が音声認識を行い、第1の実施例と同様に一部音声信号を合成音声信号に置換し、一部が合成音声信号に置換された音声信号を符号化し、符号化された音声データを単独または文字データや画像データなどと一緒にデータ通信制御部7およびデータ通信ネットワークを介して他の端末装置やサーバへ送信する。また、他の端末装置やサーバからの音声データをデータ通信制御部7により受信すると、音声出力部9によりアナログの音声信号に変換し、スピーカに出力する。こうして、この実施例によれば、データ通信手段を用いた音声による対話や討論、投稿などにおいても第1の実施例と同様の効果を得ることができる。以上、図1に示した音声処理システムの場合について説明したが、説明したような音声処理方法に従ってプログラミングしたプログラムを例えば着脱可能な記憶媒体に記憶し、その記憶媒体をこれまで本発明の音声処理を行えなかったパーソナルコンピュータなど情報処理装置に装着することにより、その情報処理装置においても本発明の音声処理を行うことができる。

【0008】

6

【発明の効果】以上説明したように、本発明によれば、請求項1記載の発明では、話者の発した音声信号について音声認識が行われ、認識された少なくとも一部の音声データが記憶されている音声データに変換され、変換された音声データから合成音声信号が生成され、前記音声信号中の対応する部分が生成された合成音声信号に置換されるので、話者の特有な言い回しや独特な表現方法、口癖など様々な特徴を含む言葉を特徴のない一般的な言葉に変換することができ、したがって、話者が誰であるかわからなくできる。また、請求項2記載の発明では、請求項1記載の発明において、変換前の単語と変換後の単語の対が登録しておかれ、認識された音声信号中の単語が変換前の単語として登録されていた場合、登録されていた前記単語が対応する変換後の単語に変換されるので、話者の特有な言い回しや独特な表現方法、口癖など様々な特徴を含む言葉を特徴のない一般的な言葉に対応づけて登録しておくことにより、請求項1記載の発明の効果を容易に実現することができる。また、請求項3記載の発明では、請求項1または請求項2記載の発明において、置換された合成音声信号を含む音声信号が公衆電話回線へ送出されるので、例えば公衆電話回線を介した匿名の討論などを行なう際に、請求項1記載の発明の効果を實現することができる。また、請求項4記載の発明では、請求項1または請求項2記載の発明において、置換された合成音声信号を含む音声信号がデータ通信ネットワークへ送出されるので、音声メールなどによる匿名の討論などを行なう際に、請求項1記載の発明の効果を實現することができる。また、請求項5記載の発明では、話者の発した音声信号について音声認識が行われ、認識された音声データが登録されていたならば、その音声データが対応づけて記憶されている他の音声データに変換され、変換された音声データから合成音声信号が生成され、前記音声信号中の対応する部分が生成された合成音声信号に置換されるので、請求項2記載の発明と同様の効果を得ることができる。また、請求項6記載の発明では、請求項5記載の音声処理方法を実施するためのプログラムが例えば着脱可能な記憶媒体に記憶されるので、その記憶媒体をこれまで請求項5記載の発明の音声処理を行えなかったパーソナルコンピュータなど情報処理装置に装着することにより、その情報処理装置においても請求項5記載の発明の効果を獲得することができる。

【図面の簡単な説明】

【図1】本発明の第1の実施例を示す音声処理システムの構成ブロック図である。

【図2】本発明の第1の実施例を示す音声処理システムの動作フロー図である。

【図3】本発明の第1の実施例を示す音声処理システム要部のデータ構成図である。

【図4】本発明の第2の実施例を示す音声処理システムの構成ブロック図である。

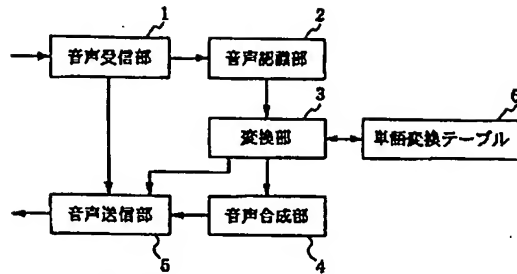
(5)

【符号の説明】

- 1 音声受信部
2 音声認識部
3 変換部
4 音声合成部

- 5 音声送信部
6 単語変換テーブル
7 データ通信制御部
8 音声入力部
9 音声出力部

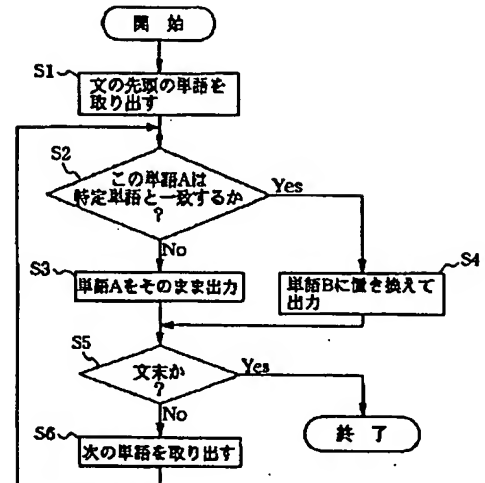
【図1】



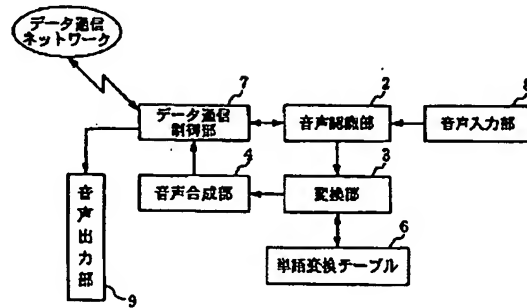
【図3】

変換前の単語	変換後の単語
シーブラブラ	シープラスプラス
シータク	タクシー
しちゃって	してしまって

【図2】



【図4】



フロントページの続き

(51) Int. Cl. 7

識別記号

F I

テマコード* (参考)

G 1 0 L 13/04

G 1 0 L 3/00

5 5 1 A 5 K 1 0 1

H 0 4 M 1/00

5 6 1 D

3/50

5 6 1 H

11/00

3 0 2

5/02

J